DOCUMENT RESUME

ED 260 595 FL 015 175

AUTHOR Powers, Stephen; And Others
TITLE An Application of the Rasch Model for Comparing the Performance of English-Speaking and Spanish-Speaking Children.
PUB DATE Aug 85
NOTE 19p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Rating; *Comparative Analysis; Ethnic Groups; Item Analysis; *Preschool Children; *School Readiness; Screening Tests; *Spanish Speaking; *Test Bias; Test Theory
IDENTIFIERS Cooperative Preschool Inventory; *English Speaking; *Rasch Model

ABSTRACT

In a study of the usefulness of the Rasch model for examining tests for possible bias, 102 native Spanish-speaking and 104 native English-speaking preschool four-year-olds in a remedial education program were administered Spanish and English versions of the Cooperative Preschool Inventory, a standardized measure of school readiness, The Rasch model of analysis was applied to the verbal and motor scales of each version. Results indicated that eight items that fit the model appeared to be improperly functioning items because on four items English-speaking pupils had an advantage over Spanish-speaking pupils, and on four, the advantage was reversed. Several discrepancies were found in the item translations and in the administration and scoring directions of the Spanish and English versions, including more complete examiner information on the English version in the form of correct responses, suggested probes, and possible answers from the examinees. In addition, the directions associated with each item in the Spanish version are given in English, requiring the examiner to translate them into Spanish before directing them to the examinee, and some of the English-to-Spanish translations allow for the change of verb tenses. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED260595

AN APPLICATION OF THE RASCH MODEL FOR COMPARING THE PERFORMANCE OF ENGLISH-SPEAKING AND SPANISH-SPEAKING CHILDREN

Stephen Powers

Tucson Unified School District

Richard L. Lopez, Jr. and Peggy Douglas

University of Arizona

August 1985

Tucson, Arizona

FL015175

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephen Powers

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

A total of 102 Spanish-speaking preschool four-year-old pupils and 104 English-speaking four-year-old pupils were individually administered the Spanish and English versions of the Cooperative Preschool Inventory (CPI). The Rasch model was applied separately to the Spanish verbal, Spanish motor, English verbal, and English motor scales of the CPI. Eight items which fit the model appeared to be improperly functioning items in the sense that on four items English-speaking pupils had an advantage over the Spanish-speaking pupils, and on four of the items, the advantage was reversed. Several differences in the items of the Spanish and English versions of the CPI are noted, as well as substantial differences between the administration and scoring directions for the two language versions.

The fidelity of translations of psychological scales has been a concern of educational researchers, school psychologists, and teachers. High quality translations allow the examination of psychological constructs in different cultures and in groups speaking different languages. Hulin, Drasgow, and Parsons (1983) summarized four types of translations: (a) the pragmatic translation where the primary purpose is to communicate accurately in the target language, (b) the aesthetic-poetic translation in which the purpose is to evoke moods, feelings, and affect in the target language, (c) the ethnographic translation in which a major aim is to maintain the cultural content of the source language, and (d) the linguistic translation which is concerned with the equivalence of meanings of both morphemes and grammatical forms of the two languages.

Numerous methods have been developed to examine tests for suspected bias. For a comprehensive review of these methods, refer to Berk (1982). Item response theory, or latent trait theory, is useful for comparing language translations because it provides evidence "whether the relation between the underlying trait and the probability of endorsing an item is identical across cultures" (Hulin et al., 1983, p. 192). This approach can be represented by an item characteristic curve of three parameters: a guessing parameter, a discrimination parameter, and a difficulty parameter. The assumptions of the Rasch model (Rasch, 1980) are (a) there is no guessing on the test, (b) all items are equally discriminating, and

(c) the test is homogeneous. In the Rasch model, the probability of a correct response to an item is a function of an examinee's ability and only one item parameter, difficulty (Ironson, 1982).

The purpose of this study was to use the Rasch model to compare the item responses of preschool children tested with either the Spanish or English versions of the Cooperative Preschool Inventory (CPI) (Caldwell, 1970a, 1974a). To the knowledge of these researchers, the Rasch model has not been applied to this inventory although this inventory has been extensively researched.

Method

Sample

The present study consisted of two independent samples of preschool four-year-old pupils enrolled in the same remedial education program in the Fall of 1982, 1983, and 1984. The first sample consisted of 102 Spanish-speaking pupils (42 boys and 60 girls). The ethnic background of this sample comprised 1 Black, 2 Native American, and 99 Hispanic children. The second sample consisted of 104 English-speaking pupils (44 boys and 60 girls). The ethnic background of the second sample was 15 Black, 10 Native American, 1 Asian, and 78 Hispanic children. Students were identified as Spanish-speaking or English-speaking by their classroom teachers based on classroom observations of the pupils for approximately one month. All pupils were enrolled in the same remedial education program in the same

large, urban school district of the Southwest. Eligibility for this program which focused on raising reading, language arts, and mathematics skills included the following criteria: (a) the child must be the sibling of an older educationally disadvantaged child, (b) at least one parent of the child lacks a high school education, (c) the child participates in a free lunch program, and (d) the child has limited proficiency in English.

Instrument

The English CPI (Caldwell, 1970a) is an individually administered English language inventory of school-readiness. A Spanish translation of the CPI (Caldwell, 1974a) is used in many programs to assess the school readiness of Hispanic pupils. The Spanish translation may be called a pragmatic translation since the primary purpose of the translation is to communicate accurately in the target language. The Spanish version of the CPI is a direct, literal translation of the English which is administered individually by a Spanish-speaking examiner. The CPI is administered in about 15 minutes and pupil responses are scored as correct or incorrect. The CPI consists of 64 items which are grouped into two subscales: (a) a verbal scale of 33 items, and (b) a motor scale of 35 items. Four items of the CPI are considered part of both the verbal and motor subscales. This instrument is designed to be a brief assessment and screening procedure for individual use with children in the age range of 3 to 6 years. It is employed variously as a screening device, a

school-readiness measure, an achievement test, and an evaluation instrument. Many school districts use the CPI to identify those individuals unprepared for traditional programs.

Previous research has supported the reliability and validity of the CPI. Powers and Medina (1984) reported alpha reliability estimates of .92 for the English CPI and .90 for the Spanish CPI. In a later study, Powers and Medina (in press) reported that the factor structure of the inventory for Spanish and English versions were similar.

Procedure

Pupils entering the preschool program were tested individually in October 1982, 1983, and 1984 with the Spanish or English CPI. These language versions were administered approximately one month after the beginning of school so that the child would become accustomed to the new surroundings and to the teacher. Further, the teacher was able to observe the students' language production in a natural setting and to determine the child's predominant language.

Rasch item difficulty estimates and person ability estimates were obtained using a microcomputer program (Powers, 1985) which utilized an unconditional maximum likelihood iterative procedure described in Wright and Stone (1979). Two primary methods for examining bias with the Rasch model were employed in this study.

They were (a) the analysis of the fit of each item to the Rash model where the item should either fit or fail to fit the model in a similar way for both groups, and (b) the comparison of the differences in difficulty parameter estimates for each item which would be estimated separately for each group (Ironson, 1982).

Results and Discussion

The assumptions of the Rasch model were first examined. Guessing was assumed to be negligible on this test because the pupils were naive four-year-old children and the test was administered individually. Discrimination was more of a concern and so point-biserial correlations were calculated for each scale. They ranged from .04 to .63 (Mdn = .36). This wide range of discrimination estimates indicated the assumption that all items were equally discriminating was not tenable. It was decided to eliminate items which did not fit the Rasch model and in this way meet the requirement of homogeneous item discrimination. The dimensionality of the latent trait space was examined using Lord's (1980) procedure. In this procedure latent roots are extracted from the item intercorrelation matrix of each scale with estimated communalities in the main diagonal. As explained by Lord (1980, p. 21): "If (1) the first root is large compared to the second and (2) the second root is not much larger than any of the others, then the items are approximately unidimensional." The first latent roots of the four scales (Spanish

verbal, Spanish motor, English verbal, and English motor) ranged from 6.63 to 8.13 ($\underline{M}$ = 7.15) and the second latent root ranged from 2.12 to 2.39 ($\underline{M}$ = 2.23). It was found that the first latent roots of each scale were triple the magnitude of the respective second latent roots. The second latent roots were, however, only slightly larger than the third and fourth latent roots. It was concluded that each scale was approximately unidimensional.

The mean square total item fit statistic (Wright and Stone, 1979) was calculated separately for the Spanish verbal, Spanish motor, English verbal, and English motor scales of the CPI. Large differences between the mean square fit statistics of the same item for two groups has been used as an indication of potential bias (Durovic, 1975; Shepard, Camilli, & Averill, 1980; Wright, Mead, & Draba, 1976). Durovic's operational definition of a large difference was that the mean square fit of one group would differ from the mean square fit of another group on the same item by 1.00 or more. This definition was adopted for the present study.

The differences between the mean square total item fit statistics for the 33 items of the Spanish and English versions of the verbal scale of the CPI were compared. Those differences ranged from -.49 to 1.81 ($\underline{M}$ = -.02, $\underline{SD}$ = .45). Only two items of the verbal scale appeared to have substantial differences. The Spanish and English versions of Item 24 differed by 1.81 and the two language versions of Item 36 differed by 1.08. The differences between the mean

square total item fit statistics of the 35 items of the Spanish and English versions of the motor CPI were compared. The differences between the mean square fit statistics on each item ranged from -.61 to 1.73 (M = .09, SD = .30). Only Item 4's difference of mean square fit statistics was 1.73 and it was the only motor item with a difference in fit statistics greater than 1.00.

Each mean square total item fit statistic was tested for significance with an F test (Wright and Stone, 1979). In order to declare that an item fit the Rasch model, a probability greater than .10 of the F ratio was required. The following items failed to fit the model, and so they were eliminated from further analysis: the Spanish verbal scale items 1, 24, 36, and 40; the English verbal scale items 1, 38, 40, 48, and 57; the Spanish motor scale items 12, 13, 15, 18, and 28; and the English motor scale items 12, 13, 18, 28, and 47. These items were eliminated from further analysis. The results of the misfit analysis were corroborated by the point-biserial correlations because most of the items rejected for not fitting the Rasch model had small, discrepant point-biserials. The mean square fit of Item 4 of the motor scale approached being classified as misfit with a $p < .12$, but because it did not reach the critical F ratio, it was retained for further analysis.

The item difficulty estimates of those items which fit both the English and Spanish versions of the CPI were compared. To place the item difficulties on the same scale, the mean differences of the item difficulties of the verbal or the motor scales was added to each item difficulty of the English item difficulties as a linking

constant. (Wright & Stone, 1979; Johnson, Parra, & Anderson, 1985). With both groups on the same scale, the difference between the item difficulties were standardized by dividing by the standard error of the difference between two item difficulties as described by Ironson (1982). The z-scores were compared with the normal curve deviate $z = 2.58$, $p < .01$ to determine if differences were large enough to suggest potentially biased items. This conservative critical value was adopted because of the multiple comparisons involved. Items with substantial differences between standardized item difficulty estimates are presented in Table 1.

Insert Table 1 about here

A positive z value indicates an item on the English test which is more likely to be answered correctly by the pupils. A negative z value indicates an item on the Spanish test is more likely to be answered correctly. Four of the items appear to favor the English-speaking pupils (Items 2, 4, 19, and 23) and four items appear to favor the Spanish-speaking pupils (Items 25, 27, 33, and 34). Probabilities of answering an item correctly given the ability parameter is zero, that is in the middle of the ability scale, are also given in Table 1. For example, the probability that a Spanish-speaking pupil will answer Item 19 correctly is .42 compared with an English-speaking pupil's probability of answering the same item correctly which is .74

Item bias methodology has been used in this study to examine the performance of English-speaking and Spanish-speaking pupils. Therefore, the items in Table 1 should be considered as possibly improperly functioning items or possibly biased items. Differences in the performance of Spanish-speaking and English-speaking children may be due to some subtle differences in the administration of the test, the surroundings of the test or numerous other factors. Further, it should be noted that the Spanish Item 4 of the motor scale was not a good fit to the Rasch model. This poor fit probably contributed to the large difficulty parameter estimate of 1.86 which in turn resulted in a large discrepancy between it and the English item.

The seven significant differences ($p < .01$) between Spanish and English items on the verbal scale suggest that there are some language or cultural differences contributing to these differences. As in previous research, it is often difficult or impossible to corroborate statistical findings in item bias research with judgmental findings. Too often an examination of the actual items fails to uncover reasons for the differential performance of examinees.

What is often elusive is the item x culture interaction which may affect student performance on test items. Since culture is carried and transmitted by language, it is often found that students from the same ethnocultural background who speak the language of the culture, also have deep roots in that culture. Also, it has been found that the acculturation process is facilitated greatly

by the degree by which one learns the language of the second culture because it is the language which conveys the new culture.

English-speaking pupils' probability of success which exceeded the probability of success of the Spanish-speaking pupils concerned the ability to tell one's age (Item 2), to show one's shoulder (Item 4), to know who to go to when sick (Item 19), what to do to read something (Item 23). The Spanish-speaking child's probability of success exceeded English-speaking children's probability of success on the following items: knowing what a mother does (Item 25), knowing what the teacher does (Item 27), knowing how many hands one has (Item 33), and knowing how many wheels on a bicycle (Item 34). Among these items, it appears that the mother's role and the teacher's role and function is more clear to the Spanish-speaking child. However, such suggested explanations must be confirmed or not with further inquiry into the differences in children's performance on translations of tests.

This study has found that eight items of an inventory may be improperly functioning. Of the eight identified items, English-speaking pupils had an advantage over the Spanish-speaking children. On the other half of the identified items, the Spanish-speaking pupils had the advantage. It has been suggested that the reasons for some of the differences may be due to cultural factors. Overall, when total scores are employed, Spanish-speaking or English-speaking advantages may be blurred or erased.

The Spanish and English versions of the CPI differ most noticeably in the administration and scoring procedures, although even the translations of questions differ. One interesting difference between the two versions is that in the Spanish version, the question of the examiner to the examinee is in Spanish but the directions to the examiner are in English. In the Spanish version, the probe which the examiner uses to elicit more information from the examinee is in English which means that there could be a variety of translations of the probe from English to Spanish. In the English version, on the other hand, the probe is often enclosed in quotes indicating that the exact wording should be used.

Another important difference between the two translations is that the English version provides the examiner with more information than does the Spanish version. A good example of this is Item 22 of the English version which provides the examiner with three of the possible correct answers. Item 22 of the English version also provides the examiner with an example of an ambiguous answer and suggests that the examiner should use a probe. Further, on the English version the examiner is provided with an example of what a correct answer to the probe might be. Item 22 of the Spanish version provides the examiner with only the question to ask. In the Spanish version, the examiner is not given any of the information about correct answers, probes and possible answers that are provided in the English version of Item 22.

Some differences in the translations were also found. Item 19 is asked in English (Caldwell, 1970b, p. 6) in the subjunctive and the conditional "If you were. . . would you. . . "The corresponding Spanish item asks in the present and future, "Si estas. . ., vas a ver?" (Caldwell, 1974b, p. 7). Other differences in the translations occur in Item 23 where the English version is in the past subjunctive and the Spanish is in the present tense.

In summary, in the English version of the CPI more information is provided to the examiner in the form of correct responses, suggested probes and possible answers from the examinee. Further, because the directions associated with each item in the Spanish version are in English, the examiner must translate some statements into Spanish before directing them to the examinee. Finally, some of the translations from English to Spanish allow for the change of verb tenses.

Care should be taken in the testing of pupils who speak a language other than English. The Spanish CPI appears on a casual inspection to be an equivalent version of the English CPI. On closer inspection there are important differences. Educational researchers, school psychologists and teachers should compare Spanish and English and their test administration procedures so that correct answers, the probes and the answers and scoring of the two versions can be standardized and comparable.

References

Berk, R. A. (1982). Handbook of methods for detecting test bias. Baltimore, MD: Johns Hopkins.

Caldwell, B. M. (1970a). Cooperative Preschool Inventory (rev. ed.). Menlo Park, CA: Addison-Wesley.

Caldwell, B. M. (1970b). Cooperative Preschool Inventory: Directions for administering and scoring (rev. ed.). Menlo Park, CA: Addison-Wesley.

Caldwell, B. M. (1974a). Cooperative Preschool Inventory (Spanish ed.). Menlo Park, CA: Addison-Wesley.

Caldwell, B. M. (1974b). Cooperative Preschool Inventory: Directions for administering and scoring (Spanish ed.). Menlo Park, CA: Addison-Wesley.

Durovic, J. J. (1975, October). Test bias: An objective definition for test items. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory: Application to psychological measurement. Homewood, IL: Dow Jones-Irwin.

Ironson, G. H. (1982). Use of chi-square and latent trait approaches for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 117-160). Baltimore, MD: Johns Hopkins.

Johnson, D. M., Parra, E. B., & Anderson, D. O. (1985, April). A study of item bias on six bilingual measures of preschool development. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum

Powers, S. (1985). RMAX: A microcomputer program for Rasch calibration of test items. Educational and Psychological Measurement, 45, 187-188.

Powers, S., & Medina, M., Jr. (1984). Reliability and validity of the Cooperative Preschool Inventory for English- and Spanish-speaking Hispanic children. Educational and Psychological Measurement, 44, 963-966.

Powers, S., & Medina, M., Jr. (in press). Factorial validity of the Cooperative Preschool Inventory for English- and Spanish-speaking Hispanic children. Journal of Psychology.

Rasch, G. (1980). Probability models for some intelligence and attainment tests. Chicago: University of Chicago.

Shepard, L. A., Camilli, G., & Averill, M. (1980, April). Comparison of six procedures for detecting test item bias using both internal and external ability criteria. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston.

Wright, B. D., Mead, R. J., & Draba, R. (1976). Detecting and correcting test item bias with a logistic response model (Research Memorandum No. 22). Chicago: Statistical Laboratory, Department of Education, University of Chicago.

Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago: MESA.

Table 1

Items Showing Substantial Differences Between Spanish and English Versions of the CPI

	Spanish			English			
Item	d	SE	p	d	SE	p	z
Verbal							
2	-.06	.24	.51	-1.22	.25	.77	3.35**
19	.34	.24	.42	-1.05	.24	.74	4.01**
23	.59	.26	.36	- .41	.23	.60	2.88*
25	-1.44	.24	.81	- .41	.23	.60	-3.10*
27	.17	.24	.46	.25	.24	.22	-3.18*
33	- .80	.23	.69	.22	.22	.45	-3.20*
34	- .75	.23	.68	.22	.22	.45	-3.05*
Motor							
4	1.86	.35	.13	.03	.22	.49	4.43**

*$z = 2.58$, $p < .01$

**$z = 3.29$, $p < .001$

Author Information

Stephen Powers, PhD, is a Researcher with the Tucson Unified School District and an Adjunct Professor in the Divisions of Educational Administration & Policy and Educational Foundations. Address: Legal and Research Services, Tucson Unified School District, P. O. Box 40400, Tucson, Arizona 85717.

Richard L. Lopez, Jr., EdD, is an Assistant Professor of Elementary Education and President of the Arizona Association for Bilingual Education (1985-86). Address: Division of Learning and Instruction, College of Education, University of Arizona, Tucson, Arizona 85721.

Peggy Douglas, PhD, is Director of the Precollege Program for Gifted and Talented Students. Address: Precollege Program for Gifted and Talented Students, College of Education, University of Arizona, Tucson, Arizona 85721.